The sample autocorrelation function of non-linear time series

Basrak, Bojan

Summary

When studying a real-life time series, it is frequently reasonable to assume, possibly after a suitable transformation, that the data come from a stationary time series (X_t). This means that the finite-dimensional distributions of this sequence are invariant under shifts of time. Various stationary time series models have been studied in detail in the literature. A standard assumption is that the time series is Gaussian or, more generally, that it has a probability distribution with light tails, in the sense that $P(|X_t| > x)$ decays to zero at least exponentially.

There is plenty of empirical (statistical) evidence that various real-life data sets, including many from insurance, finance and telecommunications, do not come from models with light tails. On the contrary, those data sets often exhibit tail behaviour which is well modelled by some power law. In applications, the question as to how fat the tails are is of considerable importance. In the financial or actuarial context, for example, the heaviness of the tails tells one about the risk one incurs, i.e. about the severity of losses one may encounter. In the context of telecommunications the heaviness of the tails of the ON- or OFF- periods of computers or of the lengths of file sizes sent from one source to another provides an indication of the efficiency of large scale stochastic networks, such as the Internet.

The heaviness of the tails is flexibly described by regular variation. It essentially means that the distributional tail $P(|X_t| > x)$ behaves like a power law $x^{-\alpha}$ for large values x en for some $\alpha > 0$. In particular, the moments $E|X_t|^h$ do not exist for $h > \alpha$, and the smaller α the heavier the tail. The notion of regular variation refers to the tail of the distribution of X_t; it does not address the modelling of the distribution in its center.

Time series analysis mostly deals with modelling of dependence in stationary random sequences. Dependence is visible in almost all real-life data sets, e.g. in financial return series where very large and very small values tend to occur in clusters over time. The standard tools of classical time series analysis are, in most parts, oriented towards second order properties, involving the analysis of correlations, covariances or mixing properties. They are much less useful if, like in the case of financial returns, heavy tails appear. That is why different tools are needed to describe "dependence in the tails". For that purpose we have chosen the

concept of *multivariate regular variation* as a suitable tool. It is an extension of power law behaviour of the tails to the multivariate case. Writing $\mathbf{X}$ to denote a d-dimensional random vector, it means that the probability $P(|\mathbf{X}| > x, \mathbf{X}/|\mathbf{X}| \in B)$ is approximated for large x by $x^{-\alpha} P(\boldsymbol{\theta} \in B)$, where B is a subset of the unit sphere of $\mathbb{R}^d$ and $|\cdot|$ denotes the Euclidean norm. The dependence between the very large values of the components of the vector $\mathbf{X}$ is completely described by the distribution of the random vector $\boldsymbol{\theta}$, the so-called *spectral measure*.

In this thesis we study various models of non-linear time series with heavy tails. Among them are bilinear processes and members of the ARCH family (autoregressive conditionally heteroscedastic processes). The latter were introduced by Engle (1982) and Bollerslev (1986) in order to model the specific form of dependence and erratic behaviour which is typical for financial return series.

We embed these standard models in a wide class of non-linear models generated by stochastic recurrence equations of the form $X_t = A_t X_{t-1} + B_t$, where (A_t, B_t) is an i.i.d. sequence, A_t is possibly matrix valued and B_t is a random vector. Obviously, these models can be understood as random coefficient autoregressive processes. The theory of these stochastic recurrence equations is sufficiently well understood. In particular, classical work gives sufficient conditions for the existence of stationary solutions (X_t), while excellent work of Kesten (1973) and Goldie (1991) tells us that the finite dimensional distributions of (X_t) are multivariate regularly varying, under very general conditions on the distribution of (A_t, B_t). In particular, we obtain conditions for the stationarity of bilinear and general ARCH processes and show that their finite-dimensional distributions are regularly varying. The latter result, to some extent, explains the presence of clusters of high/low level exceedances in real-life financial data and their unusual erratic behaviour.

The main emphasis of this thesis is on the asymptotic behaviour of the sample autocorrelations and sample autocovariances of the non-linear time series models studied. In classical time series analysis the sample autocorrelation function is one of the main tools to describe second order dependence in the series; it is the basis for model fitting and prediction purposes.

In contrast to linear time series models (such as the standard ARMA models), the sample autocorrelations of solutions to stochastic recurrence equations can display a very peculiar behaviour. In particular, if the index of regular variation α of these series is very small, the sample autocorrelation can have random limits. For moderately large α-values, the rate of convergence of the sample autocorrelations to their deterministic counterparts, the autocorrelations, can be very slow. Only if α is sufficiently large, one can use the standard asymptotic central limit theory with $\sqrt{n}$-rates and normal limits. In all the other cases, one has unusually wide asymptotic confidence bands and non-standard limits involving *stable distributions*.

A conclusion we may draw from the theory is that the sample autocorrelations of non-linear time series, in particular bilinear, ARCH and related processes, are not reliable statistical instruments for measuring dependence when the underlying series has very heavy tails. For heavy-tailed data the spectral measure of multivariate regular variation is an alternative tool to describe the dependence in the tails. In this thesis, we propagate multivariate regular variation as a probabilistic concept which describes heavy tails *together with* dependence.

Finally, we indicate under certain conditions, how one can estimate the tail parameter α and the spectral measure $P(\boldsymbol{\theta} \in \cdot\,)$. We show consistency of these estimators and apply them to simulated and real-life time series.